

عنوان مقاله:

یک روش آماری مبتنی بر پیکره برای جداسازی واژه‌های به هم چسبیده

محل انتشار:

دومین کنگره مشترک سیستم‌های فازی و هوشمند ایران (سال: 1387)

تعداد صفحات اصل مقاله: 4

نویسندگان:

محسن مشکی - دانشکده مهندسی کامپیوتر دانشگاه علم و صنعت ایران

مرتضی آنالویی

خلاصه مقاله:

در این مقاله، یک روش آماری مبتنی بر پیکره برای شناسایی واژه‌های به هم چسبیده معرفی شد. این روش، با استفاده از معیار احتمال شرطی متقارن تعیین میکند که آیا یک عنصر متنی، یک واژه‌ی کامل است یا مجموعه‌ای از واژه‌های به هم چسبیده است. روش پیشنهادی روی یک مجموعه‌ی 18000 تایی از اخبار ورزشی مورد آزمایش قرار گرفت و نتایج قابل قبولی را نشان داد. به ازای هر دوآستانه‌ی 0,001 و 0,0001 تعداد نمونه‌هایی که به درستی از هم جدا شدند به کل نمونه‌های جدا شده نسبت قابل قبولی (بیش از 80 درصد در مورد نمونه‌های جدا شده به دو بخش) را نشان داد.

کلمات کلیدی:

پیکره‌ی متنی، تعیین مرز واژه‌ها، باهم‌آیی، احتمال شرطی متقارن

لینک ثابت مقاله در پایگاه سیویلیکا:

<https://civilica.com/doc/204044>

